Appendix B
Notation and Formulae for
Separate Ratio and Regression Estimators

Invoice Value:

Invoice Value	x_i	The ith invoice item transaction amount in a sample ***or*** population, where x_8 is the invoice value from the eighth sampling unit.
Invoice Value Total (Population)	X	Invoice total from a population of ***N*** units: $X = \sum_{i=1}^{N} x_i = x_1 + x_2 + x_3 + \ldots + x_N$
Mean Invoice Value (Population)	$\overline{X}$	$\overline{X} = \frac{\sum_{i=1}^{N} x_i}{N} = \frac{x_1 + x_2 + x_3 + \ldots + x_N}{N}$
Mean Invoice Value (Sample)	$\overline{x}$	$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{x_1 + x_2 + x_3 + \ldots + x_n}{n}$
Estimated Total Invoice Value (Population)	$\hat{X}$	From an estimator (see mean, difference, ratio, or regression estimators)

Taxable (Tax) Error:

Taxable Error Value	y_i	The ith taxable error value in a sample ***or*** population, where y_8 is the taxable error value from the eighth sampling unit.
Taxable Error Total (Population)	Y	Total taxable error from a population of ***N*** units: $Y = \sum_{i=1}^{N} y_i = y_1 + y_2 + y_3 + \ldots + y_N$ *(Cochran page 20)*
Mean Taxable Error (Population)	$\overline{Y}$	$\overline{Y} = \frac{\sum_{i=1}^{N} y_i}{N} = \frac{y_1 + y_2 + y_3 + \ldots + y_N}{N}$ *(Cochran page 20)*
Mean Taxable Error (Sample)	$\overline{y}$	$\overline{y} = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{y_1 + y_2 + y_3 + \ldots + y_n}{n}$ *(Cochran page 20)*
Estimated Total Taxable Error (Population)	$\hat{Y}$	From an estimator (see mean, difference, ratio, or regression estimators)

General Stratified Notation:

Number of Strata Sampled	L	Count of the number strata sampled, ignoring any stratum detailed 100% or otherwise not sampled. *(Cochran page 89)*
Stratum	h	Subscript h, the hth stratum out of L strata. *(Cochran page 90)*
Item	i	Subscript i, the ith item, where $x_{h,i}$ is the ith invoice value in the hth stratum. *(Cochran page 90*
Stratum Count	N_h	The number of units in the hth stratum. *(Cochran page 90*
Population Count	N	$N_1 + N_2 + ... + N_L$ or $\sum_{h=1}^{L} N_h$ *(Cochran page 89)*
Sample Count (by Stratum)	n_h	Sample size in the hth stratum *(Cochran page 90*
Sample Count (All Strata)	n	$n_1 + n_2 + ... + n_L$ or $\sum_{h=1}^{L} n_h$ *(Cochran page 89)*
Invoice Value	$x_{h,i}$	$x_{3,45}$ is the invoice value from the 45th sampling unit in the 3rd stratum (or sample from the third stratum).
Total Invoice Value in the hth Stratum (by Stratum)	X_h	$x_{h,1} + x_{h,2} + ... + x_{h,N_h}$ or $\sum_{i=1}^{N_h} x_{h,i}$ Total Invoice Value in the hth stratum
Total Invoice Value (All Strata)	X	$X_1 + X_2 + ... + X_L$ or $\sum_{h=1}^{L} X_h$
Estimated Total Invoice Value in the hth Stratum	$\hat{X}_h$	From an estimator
Estimated Total Invoice Value (All Strata)	$\hat{X}_{st}$	Estimated Total Invoice value from an estimator (Overall, or across all sampled strata)
Sampling Fraction in the hth Stratum	f_h	n_h / N_h *(Cochran page 90)*
Finite Population Correction in the hth Stratum	fpc_h	$\sqrt{1 - f_h}$ *(Cochran page 24)*
Taxable Error Value	$y_{h,i}$	$y_{3,45}$ is the taxable error value from the 45th sampling unit in the 3rd stratum (or sample from the third stratum). *(Cochran page 90)*
Total Taxable Error Value in the hth Stratum (by Stratum)	Y_h	$y_{h,1} + y_{h,2} + ... + y_{h,N_h}$ or $\sum_{i=1}^{N_h} y_{h,i}$ Total Taxable Error Value in the hth stratum

Total Taxable Error Value (All Strata)	Y	$Y_1 + Y_2 + \ldots + Y_L$ or $\sum_{h=1}^{L} Y_h$
Estimated Total Taxable Error Value in the *h*th Stratum (by Stratum)	$\hat{Y}_h$	Estimated Total Taxable Error Value in the hth stratum (from an estimator)
Estimated Total Taxable Error Value (All Strata)	$\hat{Y}_{st}$	Estimated Total Error Value from an estimator (Overall, or across all sampled strata)
Mean Taxable Error Value from the Sample in the *h*th Stratum (by Stratum)	$\bar{y}_h$	$\frac{\sum_{i=1}^{n_h} y_{h,i}}{n_h}$ *(Cochran page 90*
Variance (Taxable Error Values in the hth Stratum)	$s_{y_h}^2$	$\frac{\sum_{i=1}^{n_h} (y_{h,i} - \bar{y}_h)^2}{n_h - 1}$ or $\frac{n_h \sum_{i=1}^{n_h} y_{h,i}^2 - (\sum_{i=1}^{n_h} y_{h,i})^2}{n_h (n_h - 1)}$
Standard Deviation (Taxable Error Values in the hth Stratum)	s_{y_h}	$\sqrt{s_{y_h}^2}$
Standard Error of the *h*th Stratum (Population Total Taxable Error Value)	$s_{\hat{Y}_h}$	$\frac{N_h s_{y_h} fpc_h}{\sqrt{n_h}}$ or $\sqrt{\frac{s_{y_h}^2 N_h (N_h - n_h)}{n_h}}$
Overall Standard Error of the Estimated Total Taxable Error Value	$s_{\hat{Y}_{st}}$	$\sqrt{\sum_{h=1}^{L} s_{\hat{Y}_h}^2}$
Effective Degrees of Freedom (Stratified Sampling)	n_e	$\frac{s_{\hat{Y}_{st}}^4}{\sum_{h=1}^{L} \left(s_{\hat{Y}_h}^4 / (n_h - 1) \right)}$ *(Cochran 5.16)*

Separate Ratio Estimator:

Stratified ratio estimation of ***Y*** can be done in two ways, with *separate* or *combined* methods (see Cochran, pages 164-167). Donald Roberts recommends the combined method for use in auditing. On page 166, Cochran states that, "the combined estimate is much less subject to risk of bias than the separate estimate." In both of these sources, advice is given when separate estimation can be considered.

Separate Ratio Estimator	***Rs***	Subscript ***Rs***
Separate Ratio in the *h*th Stratum	R_{S_h}	$\frac{\sum_{i=1}^{N_n} y_{h,i}}{\sum_{i=1}^{N_h} x_{h,i}}$ *(From the population - unknown)*
Estimated Separate Ratio in the *h*th Stratum	$\hat{R}_{S_h}$	$\frac{\sum_{i=1}^{n_n} y_{h,i}}{\sum_{i=1}^{n_h} x_{h,i}}$ *(From the sample)*
Estimated Total Taxable Error Value in the *h*th Stratum	$\hat{Y}_{Rs_h}$	$X_h \hat{R}_{S_h}$
Estimated Total Taxable Error Value (Separate Ratio Estimation)	$\hat{Y}_{Rs}$	$\sum_{h=1}^{L} \hat{Y}_{Rs_h}$
Standard Deviation of the Taxable Error Values in the *h*th Stratum (Separate Ratio Estimation)	$s_{y_{Rs,h}}$	$\sqrt{\frac{\sum_{i=1}^{n_h}\left[y_{h,i} - \left(\hat{R}_{S_h} x_{h,i}\right)\right]^2}{n_h - 1}}$
Standard Error of the *h*th Stratum (Separate Ratio Estimation)	s_{Rs_h}	$\frac{N_h s_{y_{Rs,h}} fpc_h}{\sqrt{n_h}}$ or $\sqrt{\frac{s_{y_{Rs,h}}{}^2 N_h (N_h - n_h)}{n_h}}$
Overall Standard Error of the Estimated Total Taxable Error (Separate Ratio Estimation)	$s_{\hat{Y}_{Rs}}$	$\sqrt{\sum_{h=1}^{L} s_{Rs_h}}$
	g_h	$\frac{N_h (N_h - n_h)}{n_h}$
Effective Degrees of Freedom (Stratified Sampling)	n_e	$\frac{s_{\hat{Y}_{Rs}}^4}{\sum_{h=1}^{L}\left(s_{Rs_h}^4 / (n_h - 1)\right)}$ or $\frac{\left(\sum_{h=1}^{L} g_h s_{y_{Rs,h}}^2\right)^2}{\sum_{h=1}^{L} \frac{g_h^2 s_{y_{Rs,h}}^4}{n_h - 1}}$ *(Cochran 5.16)*
Precision Amount (Separate Ratio Estimator)		$t_{n_e} s_{\hat{Y}_{Rs}}$
Confidence Interval (Separate Ratio Estimation)		$\hat{Y}_{Rs} \pm \left(t_{n_e} s_{\hat{Y}_{Rs}}\right)$

Separate Regression Estimator:

Separate Regression Estimator	***Gs***	Subscript ***Gs***
Separate Regression Coefficient in the *h*th Stratum	b_{Gs_h}	$\frac{\sum_{i=1}^{n_h}\left[\left(x_{h,i}-\bar{x}_h\right)\left(y_{h,i}-\bar{y}_h\right)\right]}{\sum_{i=1}^{n_h}\left(x_{h,i}-\bar{x}_h\right)^2}$
		Excel formula: "=Slope(range of $\boldsymbol{y_{h,i}}$ values, range containing $\boldsymbol{x_{h,i}}$ values)"
Estimated Total Taxable Error Value in the *h*th Stratum (Separate Regression Estimation)	$\hat{Y}_{Gs_h}$	$\left(N_h\overline{y_h}\right)+\left(b_{Gs_h}\left(X_h-\left(N_h\bar{x}_h\right)\right)\right)$
Estimated Total Taxable Error Value (Separate Ratio Estimation)	$\hat{Y}_{Gs}$	$\sum_{h=1}^{L}\hat{Y}_{Gs_h}$
Standard Deviation of the Taxable Error Values in the *h*th Stratum (Separate Regression Estimation)	$s_{y_{Gs,h}}$	$\sqrt{\frac{\sum_{i=1}^{n_h}\left(\left(y_{h,i}-\bar{y}_h\right)-b_{Gs_h}\left(x_{h,i}-\bar{x}_h\right)\right)^2}{n_h-2}}$
Standard Error of the *h*th Stratum (Separate Regression Estimation)	s_{Gs_h}	$\frac{N_h s_{y_{Gs,h}} fpc_h}{\sqrt{n_h}}$ or $\sqrt{\frac{s_{y_{Gs,h}}{}^2 N_h(N_h-n_h)}{n_h}}$
Overall Standard Error of the Estimated Total Taxable Error (Separate Regression Estimation)	$s_{\hat{Y}_{Gs}}$	$\sqrt{\sum_{h=1}^{L}s_{Gs_h}}$
	g_h	$\frac{N_h\left(N_h-n_h\right)}{n_h}$
Effective Degrees of Freedom (Stratified Sampling)	n_e	$\frac{s_{\hat{Y}_{Gs}}^4}{\sum_{h=1}^{L}\left(s_{Gs_h}^4/\left(n_h-1\right)\right)}$ or $\frac{\left(\sum_{h=1}^{L}g_h s_{y_{Gs,h}}^2\right)^2}{\sum_{h=1}^{L}\frac{g_h^2 s_{y_{Gs,h}}^4}{n_h-1}}$ *(Cochran 5.16)*
Precision Amount (Separate Regression Estimation)		$t_{n_e}s_{\hat{Y}_{Gs}}$
Confidence Interval (Separate Regression Estimation)		$\hat{Y}_{Gs}\pm\left(t_{n_e}s_{\hat{Y}_{Gs}}\right)$

Coefficients of Variation (Stratified Sampling):

The estimated standard error formulas found above are "large sample approximations" for ratio and regression estimation. Cochran, at page 153, recommends that the $CV_{\bar{x}}$ and $CV_{\bar{y}}$ both be below 10% as a working rule to insure that the samples are sufficiently large so that the formulas can be presumed valid.

The requirement generally applied is that $CV_{\hat{X}}$ always be below 10%. In using separate estimation, $CV_{\hat{X}_h}$ must be below 10% in each stratum in order to be able to use the overall estimates of total taxable error for ratio and regression estimation, $\hat{Y}_{Rs}$ and $\hat{Y}_{Gs}$.

Estimated standard deviation of the invoice amounts in the *h*th stratum	s_{x_h}	$\sqrt{\dfrac{\sum_{i=1}^{n_h}(x_{hi}-\bar{x}_h)^2}{n_h-1}}$
Coefficient of Variation of the Estimated Total Invoice Value in the *h*th stratum (Stratified Frame) (same as the CV of the sample mean invoice value, $CV(\bar{x})_h$)	$CV(\hat{X})_h$	$\dfrac{s_{x_h}\, fpc_h \big/ \sqrt{n_h}}{\bar{x}_h}$